

# What is the distance to the CMB?

Chris Clarkson<sup>1</sup>, Obinna Umeh<sup>2</sup>, Roy Maartens<sup>2,3</sup> and Ruth Durrer<sup>4</sup>

<sup>1</sup>*Astrophysics, Cosmology & Gravity Centre, and, Department of Mathematics & Applied Mathematics, University of Cape Town, Cape Town 7701, South Africa.*

<sup>2</sup>*Physics Department, University of the Western Cape, Cape Town 7535, South Africa*

<sup>3</sup>*Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, United Kingdom*

<sup>4</sup>*Département de Physique Théorique & Center for Astroparticle Physics, Université de Genève, CH-1211 Genève 4, Switzerland.*

The success of precision cosmology depends not only on accurate observations, but also on the theoretical model – which must be understood to at least the same level of precision. Subtle relativistic effects can lead to biased measurements if they are neglected. One such effect gives a systematic shift in the distance-redshift relation away from its background value, due to the non-linear relativistic conservation of total photon flux. We also show directly how this shift follows from a fully relativistic analysis of the geodesic deviation equation. We derive the expectation value of the shift using second-order perturbations about a concordance background, and show that the distance to last scattering is increased by 1%. We argue that neglecting this shift could lead to a significant bias in the background cosmological parameters, because it alters the meaning of the background model. A naive adjustment of CMB parameter estimation if this shift is really a correction to the background would raise the  $H_0$  value inferred from the CMB by 5%, potentially removing the tension with local measurements of  $H_0$ . Other CMB parameters which depend on the distance would also be shifted by  $\sim 1\sigma$  when combined with local  $H_0$  data. While our estimations rely on a simplistic analysis, they nevertheless illustrate that accurately defining the background model in terms of the expectation values of observables is critical when we aim to determine the model parameters at the sub-percent level.

## I. INTRODUCTION

Cosmology has entered a precision era. The premier cosmological dataset is the anisotropies and polarization of the cosmic microwave background (CMB). This is not only due to the highly accurate data, but also because of its simple theoretical description, which allows accurate calculations. Present CMB codes like CAMB [1] and CLASS [2] are typically 0.1% accurate and, together with contemporary data, provide a determination of basic cosmological parameters to the percent level (and substantially better for the case of curvature) [3]. A puzzling result from current CMB measurements is that the Hubble parameter  $H_0$  is significantly smaller than the value measured locally [4–6].

Parameter estimation from the CMB is extremely sensitive to the angular diameter distance  $d_A(z_*)$ , where  $z_* \simeq 1090$  is the redshift of the last scattering surface. More precisely it depends on the angular size of the sound horizon,  $\theta = r_*/d_A(z_*)$ , where  $r_*$  is the sound horizon at last scattering. The Planck collaboration [3] has reported  $\theta = (1.04131 \pm 0.00063) \times 10^{-2}$ ; hence it measures this scale with an accuracy of better than  $10^{-3}$ . The accuracy of  $r_*$  is slightly worse, about  $4.3 \times 10^{-3}$ , which is also the accuracy of  $d_A(z_*)$ . These numbers indicate that a change of  $\sim 1$  percent in  $d_A(z_*)$  is critical for parameter estimation of the CMB at the present level of accuracy.

Most calculations of the CMB anisotropies are performed within first-order perturbation theory and only CMB lensing requires a second-order analysis including

first-order angular deflection of first-order temperature anisotropies and polarization. In this work we consider the homogeneous change in the angular diameter distance due to the presence of structures in the Universe, up to second order in perturbation theory. We argue that it is critical to include this change at the present level of accuracy, as it induces a change in the mean distance to the last scattering surface which is larger than the current measurement error. Including this change could remove the tension between the values of  $H_0$  obtained by CMB observations and by local measurements. This change then affects other parameters which depend on  $H_0$ . For example,  $\omega_m = \Omega_m h^2$  is well determined by the CMB, and so a change in  $h$  implies  $\Omega_m$  moves significantly away from its naive value inferred without inclusion of this relativistic second-order correction. (Here  $h$  is defined by  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ .)

The observed angular diameter distance at observed redshift  $z_s$  in direction  $\mathbf{n}$  is

$$d_A(z_s, \mathbf{n}) = \frac{1}{1+z_s} \left[ \chi_s + \delta d_A(z_s, \mathbf{n}) + \frac{1}{2} \delta^2 d_A(z_s, \mathbf{n}) \right] = \bar{d}_A(z_s) [1 + \Delta(z_s, \mathbf{n})], \quad (1)$$

which has a perturbation  $\Delta(z_s, \mathbf{n})$  about the the background distance

$$\bar{d}_A(z_s) = \frac{\chi_s}{(1+z_s)} = \frac{1}{(1+z_s)} \int_0^{z_s} \frac{dz}{(1+z)\mathcal{H}(z)}. \quad (2)$$

Here  $\chi_s$  is the comoving distance (in the background geometry) to the source at observed redshift  $z_s$  and  $\mathcal{H}$  is the

conformal Hubble rate. The perturbation  $\Delta(z_s, \mathbf{n})$  arises from the fact that the Universe is not homogeneous and isotropic, but contains cosmic structures which induce fluctuations in the geometry. At linear order in perturbation theory, the lensing convergence  $\kappa = -\Delta$  produces no change in the mean value  $\langle d_A \rangle$  (although it does give a variance) [7]. At second order, however, nonlinear effects give a relativistic correction to the distance-redshift relation that a typical observer would expect.

This correction can be calculated from the ensemble average:

$$d_A^{\text{eff}}(z_s) = \langle d_A(z_s, \mathbf{n}) \rangle = \bar{d}_A(z_s)[1 + \langle \Delta \rangle(z_s)], \quad (3)$$

where we assume statistically isotropic Gaussian initial perturbations, so that there is no dependence on directions (all directions receive the same correction). If it is not correctly taken into account, the shift in the ‘background’ distance-redshift relation by  $\langle \Delta \rangle$  could result in a shift in the inferred cosmological parameters which appear in the distance-redshift relation. We begin by anticipating that this correction is imposed by conservation of flux, and we show rigorously in the appendices that this follows from purely geometrical arguments using the geodesic deviation equation for null geodesics.

### Flux conservation requires a shift in angular diameter distance

Gravitational lensing conserves the total photon flux  $F = L/(4\pi d_L^2)$ , where  $d_L$  is the luminosity distance. This is a fully nonlinear and general result based on the analysis of light rays in curved spacetime. In the case of the CMB, this implies that a spatial average of the flux must be conserved from one time to the next. Because the CMB is emitted when the universe is very smooth, this implies that we should expect  $\langle F \rangle \simeq \bar{F}$  where  $\bar{F}$  is the value in the background. Since  $d_L = (1 + z_s)^2 d_A$ , this implies the average of  $d_A^{-2}$  should be close to its background value at the same observed redshift. This means,  $\langle d_A^{\text{eff}}(z_s)^{-2} \rangle \simeq \bar{d}_A(z_s)^{-2}$ , where the approximate equality means there are no significant corrections, i.e., no corrections as large as  $\mathcal{O}(\delta^2)$ , where  $\delta$  is the matter over-density. We expand (1) perturbatively,

$$d_A^{-2} = \frac{(1+z)^2}{\chi^2} \left[ 1 - 2\frac{\delta d_A}{\chi} - \frac{\delta^2 d_A}{\chi} + 3\left(\frac{\delta d_A}{\chi}\right)^2 \right]. \quad (4)$$

The linear term vanishes on average by definition:  $\langle \delta d_A \rangle = 0$ . Then, flux conservation implies, on average,  $\langle \delta^2 d_A \rangle \simeq 3\langle (\delta d_A)^2 \rangle / \chi$  and consequently

$$\langle \Delta \rangle \simeq \frac{3}{2} \left\langle \left( \frac{\delta d_A}{\chi} \right)^2 \right\rangle = \frac{3}{2} \langle \kappa^2 \rangle. \quad (5)$$

where  $\kappa$  is the usual linear lensing convergence. This is actually the leading contribution to the expected change

to large distances. We prove this remarkably simple and important result in a variety of ways in several appendices. It implies that the total area of a sphere of constant redshift will be larger than in the background. Physically this is because a sphere about us in redshift space is not a sphere in real space – lensing implies that this ‘sphere’ becomes significantly crumpled in real space, and hence has a larger area. When interpreted as a shift to the background geometry, this would have important implications for the analysis of the CMB. An observed patch of the CMB sky such as a hot or cold spot of a fixed observed angular scale will correspond to a physical area which is larger than the background value, since the distance to it is larger. Effectively, it is the angular size of these hot and cold spots, combined with a theoretical model for calculating both the distance to the CMB and the sound horizon scale at last scattering, that determine many key parameters of the cosmological standard model. Consequently, we anticipate a shift in the inferred background cosmology when aggregated lensing is taken into account.

Here we quantify this shift for a flat  $\Lambda$ CDM (concordance) background, see the result plotted in Fig. 1, and we explore the potential consequences for precision cosmology. At low redshifts the change to  $\bar{d}_A$  is small ( $|\langle \Delta \rangle| \lesssim 10^{-4}$ ), negative and dominated by local effects (from coupled velocity and Sachs-Wolfe terms). Recently [8] have estimated the effect of this change on  $H_0$  and especially on its variance measured with low- $z$  data. For  $z \gtrsim 0.5$  the change becomes positive and is dominated by second-order lensing effects. It grows monotonically until last scattering, and the distance to the CMB is increased by about one percent.

If this shift is not taken into account, the distance-redshift relation for flat  $\Lambda$ CDM will appear to be that of a slightly open model ( $\Omega_K > 0$ ), due to the slight shrinking of objects relative to the background. At high redshift  $z \gtrsim 10$ , the effective distance is similar to an open  $\Lambda$ CDM model. At the CMB redshift, measurement of one value  $d_A(z_*)$  requires a change in  $H_0$  in the background model to account for this shift.

The incorporation of perturbations to the monopole of the local  $d(z)$  relation was first considered in [9, 10] using a series expansion, where it was speculated to produce large contributions from small-scale power. (We discuss the contribution of small-scale power below.) A later study in [11, 12] used ‘lightcone averaging’ to analyze the monopole of the magnitude-redshift relation at low to moderate redshift. The effect was predicted to be small, consistent with our low redshift results, but in contrast with [9, 10]. None of these works consider the impact on cosmological parameter estimation from CMB observations. Here we focus on high- $z$ , where the aggregation of lensing events is more significant. We use ensemble averaging, so that specifically considering the monopole is unnecessary when the perturbations are statistically isotropic. The approach here is significantly simpler than the results obtained in the papers cited above. The details of the calculation, including four in-

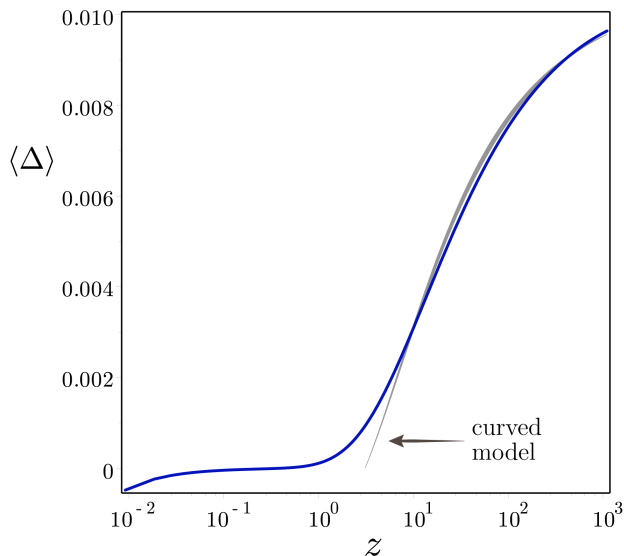


Figure 1: Fractional correction  $\langle \Delta \rangle(z)$  to the distance [see (3)] for a fiducial model  $\Omega_m = 0.3$ ,  $h = 0.68$ ,  $\omega_b^2 = 0.0222$ ,  $w = -1$  and  $n_s = 0.96$ . The correction is negative for  $z \lesssim 0.25$ , purely from the local contribution. At higher redshift the shift arises from the aggregated lensing term (6). For  $z \gtrsim 10$  the corrections grow  $\propto \chi_s^3$ , and are similar to an open  $\Lambda$ CDM model with  $\Omega_K^{\text{eff}} \approx 0.0066$  (grey ‘curved’, shown for high  $z$ ).

dependent derivations of the main formula (5), are presented in the Appendices.

## II. CORRECTION TO THE DISTANCE FROM GEODESIC DEVIATION

The second-order perturbation of  $d_A$  about a flat  $\Lambda$ CDM background in the Poisson gauge has been given in [13, 14] based on the Sachs equations (see also [15, 16] for another approach).

Assuming no anisotropic pressure and no vectors and tensors at linear order means the first-order modes are determined by a separable potential  $\Phi(\eta, \mathbf{x}) = g(\eta)\Phi_0(\mathbf{x})$ , normalized such that  $g(\eta_0) = 1$ . Here  $\eta$  is conformal time and  $\eta_0$  denotes today. The perturbation  $\Delta$  splits into local terms evaluated at either the source or observer, local times integrated terms, and integrated terms [13]. The key local terms are not relevant for us, and are discussed in Appendix F.

The key integrated terms at second-order are those with a total of 4 screen-space derivatives which do not form part of a 2-divergence in the screen space. (A pure 2-divergence gives zero contribution to the ensemble average.) In Appendix A we derive this result from the well known Sachs equation. The relevant term is, up to total divergence terms,

$$\Delta_{\text{int}}(z_s) = \frac{3}{2} \left[ \int_0^{\chi_s} \frac{d\chi}{\chi} \left( 1 - \frac{\chi}{\chi_s} \right) \Delta_2 \Phi(\chi) \right]^2 \quad (6)$$

Here  $\Delta_2$  is the angular Laplacian on the sphere of observer directions. This second-order part is the key term which contributes when an ensemble average is taken. Four different derivations of it are given in Appendix A to B.

In the full expressions given in the literature cited above, there are a variety of other terms, we have carefully checked they are sub-dominant to those which we discuss here. Any terms which form a total divergence disappear once we integrate over directions or perform an ensemble average.

The potential can be written as

$$\begin{aligned} \frac{\Phi(\chi, \mathbf{n})}{g(\eta)} &= \int \frac{d^3 k}{(2\pi)^{3/2}} \Phi_0(\mathbf{k}) e^{i\chi \mathbf{k} \cdot \mathbf{n}} = \sum_{\ell m} \Phi_{\ell m}^0(\chi) Y_{\ell m}^*(\mathbf{n}), \\ \Phi_{\ell m}^0(\chi) &= \sqrt{\frac{2}{\pi}} i^\ell \int d^3 k j_\ell(k\chi) \Phi_0(\mathbf{k}) Y_{\ell m}(\hat{\mathbf{k}}). \end{aligned} \quad (7)$$

The power spectrum of  $\Phi_0$  is

$$\langle \Phi_0(\mathbf{k}_1) \Phi_0(\mathbf{k}_2) \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_0(k) T(k)^2 \delta^3(\mathbf{k}_1 + \mathbf{k}_2), \quad (8)$$

where  $T(k)$  is the transfer function and

$$\mathcal{P}_0 = \left( \frac{3\Delta_{\mathcal{R}}(k_0)}{5g_\infty} \right)^2 \left( \frac{k}{k_0} \right)^{n_s-1}. \quad (9)$$

Here

$$g_\infty \approx \frac{1}{5} (3 + 2\Omega_m^{-0.45}), \quad \Delta_{\mathcal{R}}^2 \approx 2.4 \times 10^{-9}$$

is the squared amplitude of the primordial curvature perturbation at the pivot scale  $k_0 = 0.002 \text{ Mpc}^{-1}$ , and  $n_s - 1$  is the spectral tilt.

The expectation value of the dominant integrated contribution can be reduced to a form convenient for numerical integration (see the Appendix E for details):

$$\begin{aligned} \langle \Delta_{\text{int}} \rangle &= 6\pi \sum_{\ell=0}^{\infty} \left[ \frac{\ell(\ell+1)}{2\ell+1} \right]^2 \times \\ &\int_0^{\chi_s} d\chi \frac{(\chi_s - \chi)^2}{\chi \chi_s^2} g^2(\chi) (\mathcal{P}_0 T^2) \Big|_{k=(\ell+1/2)/\chi}. \end{aligned} \quad (10)$$

For large distances we can estimate analytically the scaling behaviour when baryons are neglected. A crude estimate of the transfer function is (adapted from [17])  $T(k = \ell/\chi) \approx 1/[1 + \alpha(\ell/\chi k_{\text{eq}})^2]$ , where  $k_{\text{eq}} \approx 0.075\Omega_m h^2 \text{ Mpc}^{-1}$  is the equality scale, and  $\alpha \sim 0.05$ . The main contribution to the sum in (10) comes from small scales with  $\ell > \chi_s k_{\text{eq}}$ . Approximately, the  $\ell$  factors out of the integral and the sum from  $\chi_s k_{\text{eq}}$  to infinity gives a factor of  $\sim 1/(4k_{\text{eq}}\chi_s)$ , assuming  $g \sim g_\infty$ . The integral becomes  $\sim (k_{\text{eq}}\chi_s)^4$ , giving  $(k_{\text{eq}}\chi_s)^3$  scaling. In fact, for this transfer function the sum over  $\ell$  and the integral can be done analytically (ignoring the  $+1/2$  in

the Limber approximation), from which we find for large  $\chi_s$ ,

$$\langle \Delta_{\text{int}} \rangle \sim 2\Delta_{\mathcal{R}}^2 (k_{\text{eq}} \chi_s)^3 \approx 0.014 \left( \frac{\Omega_m h^2}{0.14} \right)^3 \left( \frac{\chi_s}{14 \text{ Gpc}} \right)^3. \quad (11)$$

(This estimate is a reasonable approximation to the numerical result at large distances – but note it is very sensitive to  $\alpha$ ). For a standard cosmology this implies corrections around the percent level for 10 Gpc distances, making  $\langle \Delta_{\text{int}} \rangle$  the dominant part of the signal for  $z \gtrsim 1$ .

The generic behaviour of  $\langle \Delta \rangle(z)$  is shown in Fig. 1: at low  $z$  the local Doppler contribution dominates, and the amplitude is small,  $\mathcal{O}(10^{-4})$ , and negative for  $z \lesssim 0.2 - 0.3$  depending on the model. At higher  $z$ , the amplitude is positive, implying larger distances, and grows roughly linearly in  $z$  reaching near percent-level around  $z \sim 5 - 10$ , thereafter growing proportional to the volume [roughly approximated by (11)] reaching around 1% by  $z \sim 10^3$ . This is the aggregated lensing signal. A higher matter density or Hubble constant increases the amplitude of  $\langle \Delta \rangle$ , while increasing the baryon fraction or including a tilt to  $n_s < 1$  decreases it by tens of percent.

### Small-scale sensitivity

The convergence of the sum in (10) is very slow, reflecting the sensitivity to the accumulation of many small-scale lensing events. We can attempt to estimate the convergence rate analytically to determine the modes that are important. For the transfer function and approximations leading to (11), we may replace the formal sum to infinity with a cutoff at  $\ell_{\text{max}} = \chi_s k_{\text{max}}$ , implying that the sum now gives a contribution  $(1/k_{\text{eq}} - 1/k_{\text{max}})/4\chi_s$ . Consequently for percent-level accuracy we need  $k_{\text{max}} \gtrsim 100k_{\text{eq}} \sim 1 \text{ Mpc}^{-1}$ . This seems reasonable, but actually overestimates the convergence rate significantly. Including the logarithmic amplification of small-scale power in a more accurate transfer function slows the convergence so that percent accuracy is achieved for only for  $k_{\text{max}} \gtrsim 10^4 k_{\text{eq}}$ , while  $k_{\text{max}} \sim 100k_{\text{eq}}$  is only 10% accurate. This slow convergence remains also when baryons are included or small-scale power is amplified by a factor of 2 according to [19]. *This implies that modes down to 10 kpc scales in principle contribute to the aggregated lensing effect, and cannot be neglected.* Since the power spectrum is not precisely modelled on these scales, we present our results for the linear, analytical approximation to the power spectrum given in Ref. [17]. This in general underestimates the true power spectrum on small scales, however, most of the effect accumulates at high  $z$  where linear theory is accurate, so our key results should be correct.

One may ask whether we should add an ultraviolet cutoff to smooth away small scale power in  $\Phi$ . This is appropriate when binning data and calculating correlation functions, for example, as the data sample effectively

smooths over the linear matter distribution. This is not the case for the cumulative relativistic correction considered here, which is effectively a sum over all possible lensing events to give an ensemble average. Note also that, neglecting the mild time dependence of  $\Phi$  the expression in (6) is positive definite and smoothing does not reduce it.

The physically relevant dark matter free-streaming cutoff is  $\mathcal{O}(\text{pc})$ , but this does not change our results. A larger UV-cutoff – equivalent to smoothing  $\Phi$  – simply investigates a different ensemble of universes with different lensing properties on small scales, quite unlike the real universe, and underestimates the size of the effect (linear theory with no cutoff still slightly underestimates it, but not by much). For small objects like SNIa it is clear we need high  $k$ -modes as the lightrays probe such small scales [20]. For a diffuse background like the CMB it is tempting to assume that the CMB light effectively smooths the matter distribution on the scale of observation. That is true at linear order, but not at second order, where the lensing effects accumulate and do not cancel in the same way. Mathematically, if we smooth  $d_A(z, \mathbf{n})$  over a solid angle on the sky, we would eventually integrate (E2) against a window function in  $\mathbf{n}$ ; but this integral simply drops out after applying the spherical harmonic addition theorem, making no difference to  $\langle \Delta \rangle$ . This is not the same as a cutoff in  $k$ , which smooths  $\Phi$  itself, not  $d_A(z, \mathbf{n})$ . Consequently, it is important not to smooth the potential, and include power down to the smallest relevant scales.

### III. POTENTIAL IMPLICATIONS FOR THE CMB AND COSMOLOGICAL PARAMETER ESTIMATION

We now explore the potential effects on parameter estimation from the CMB under the assumption that the correction from aggregated lensing can be interpreted as a monopole shift in the distance. This assumption requires further justification, as we discuss in the Conclusions, and it is likely that our parameter estimates will be shifted when the full effects of aggregated lensing are taken into account. We leave this for further work.

If the distance to the CMB is changed by a homogeneous factor  $(1 + \Delta_*)$ ,  $d_A \rightarrow (1 + \Delta_*)d_A$ , features in the CMB are simply rescaled. More precisely, within the flat sky approximation which can be used for  $\ell > 20$ , the CMB power spectrum  $C_\ell$  changes as [22]

$$C_\ell \rightarrow (1 + \Delta_*)^2 C_{(1+\Delta_*)\ell}. \quad (12)$$

We find that the shift in the distance to the CMB is well approximated by

$$\Delta_* = 0.63 \Omega_m^{1.1 f_b + 1.8} h^{2.5 f_b + 3} n_s^5 e^{-0.28 \Omega_b / 0.0222} \quad (13)$$

where the baryon fraction is  $f_b = \Omega_b / \Omega_m$ . This is accurate to about 1% in the vicinity of the Planck parameter values, and 10% accurate well outside that range.

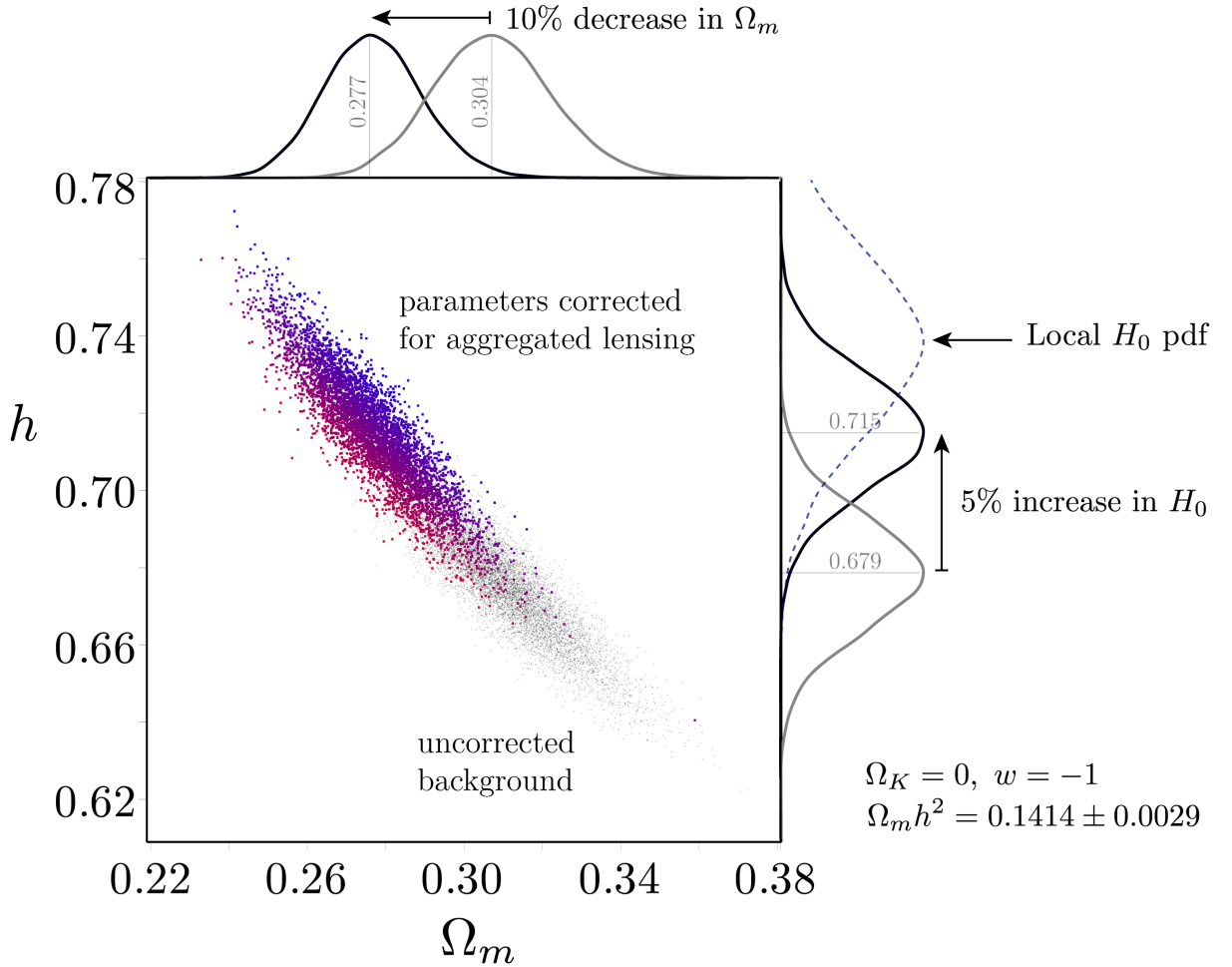


Figure 2: We show the distribution of values for  $h$  and  $\Omega_m$  from the Planck data when using only the background angular diameter distance (black dots) and when including the correction  $\Delta_*$  (red to blue dots, coloured according to  $d_A(z_*)$ , with red indicating larger distances). At the boundary we also show the one dimensional probability distribution function of the corresponding variable. The resulting Hubble parameter is now in agreement with local measurements – the dotted PDF is adapted from Riess et al. [23].

Using the Planck parameters for  $\Omega_m h^2 \equiv \omega_m$ ,  $h$ ,  $\Omega_b h^2 \equiv \omega_b$  and  $n_s$  (as independent Gaussian distributed values from ‘Planck+Lensing’, Table 2 in [3], and setting  $z_* \approx 1090$ ) we find

$$\langle \Delta \rangle(z_*) \equiv \Delta_* = 0.010285 \pm 0.00074. \quad (14)$$

Hence aggregated lensing leads to a change of about 1% in the distance to the CMB which is extremely important for CMB parameter estimation. This corresponds to  $\sim 2.5\sigma$  error on the Planck estimation of the distance, so we anticipate that background parameters dependent on this distance shift by about  $2.5\sigma$ . (Of course the Planck parameters are not independent but to some extent correlated, but this does not modify the mean value of  $\Delta_*$ , only its error bar.) The CMB measures  $d_A(z_*)$  precisely, so we must adjust  $\bar{d}_A(z_*)$  to account for  $\Delta_*$ .

In Fig. 2 we compare the values of  $h$  and  $\Omega_m$  inferred without correction with those obtained including aggregated lensing. For this we use the Planck values [3]

$\omega_m = 0.1414 \pm 0.0029$ , and  $h = 0.679 \pm 0.015$  giving the measured distance  $(1+z_*)d_A(z_*) = 13.98 \pm 0.13 \text{ Gpc}$  (we ignore the small correction from radiation), and we take  $\omega_b = 0.02217 \pm 0.00033$ ,  $n_s = 0.9635 \pm 0.0094$ . We treat these as independent Gaussian distributed variables. The values of  $\omega_m$ ,  $\omega_b$  and  $n_s$  are determined by the physics of last scattering only, and do not depend on the distance – but  $h$  does. From  $d_A(z_*) = \bar{d}_A(z_*)(1 + \Delta_*)$  we then infer the background parameters  $h$ ,  $\Omega_m$ . For a flat  $\Lambda$ CDM model we find:

$$h = 0.715 \pm 0.0153, \quad \Omega_m = 0.277 \pm 0.0125 \quad (15)$$

For  $h$  this is now in good agreement with Riess et al. [23],  $h = h_R = 0.738 \pm 0.024$ , removing all tension.

At first one might be surprised that a 1% shift in  $d_A$  yields such a large (about 5%) shift in  $h$  and  $\Omega_m$ . However, a short calculation shows that this is to be expected:

at fixed  $\omega_m$ , neglecting curvature and radiation we have

$$d_A(z_*) = \frac{L_0}{1+z_*} \int_0^{z_*} \frac{dz}{[\omega_m(1+z)^3 - \omega_m + h^2]^{1/2}}$$

$$\frac{\partial d_A}{\partial h}(z_*) = \frac{-hL_0}{1+z_*} \int_0^{z_*} \frac{dz}{[\omega_m(1+z)^3 - \omega_m + h^2]^{3/2}}.$$

where  $L_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1} \simeq 3000 \text{ Mpc}$ . If we want to absorb a change in  $d_A$  as a change in the background value of  $h$  we use the fact that the background value of the distance needs to be reduced by  $\Delta$  from its observed value:

$$d_A^{\text{obs}} = \bar{d}_A(\bar{h})(1 + \Delta) \quad \text{implies} \quad (16)$$

$$\bar{d}_A(\bar{h}) = d_A^{\text{obs}}(1 - \Delta) \quad (17)$$

$$= \bar{d}_A(h) + \frac{\partial d_A}{\partial h} h \Delta_h \quad (18)$$

we require a change in  $h$ ,  $h = \bar{h}(1 + \Delta_h)$  given by

$$\Delta_h = -\frac{d_A}{(\partial d_A / \partial h) h} \Delta \simeq +5\Delta. \quad (19)$$

For the last  $\simeq$  sign we have set  $z_* = 1090$ ,  $\omega_m = 0.14$  and  $h = 0.72$ . Hence  $h$  is very sensitive to  $d_A$ . A 1% change in  $d_A$  does necessitate a 5% increase in  $h$ , and consequently a 10% decrease in  $\Omega_m$ , as we have observed in our analysis.

In a full joint analysis not only  $h$  and  $\Omega_m$  but also other parameters will shift. As an example we consider curvature [27] or evolving dark energy with a constant equation of state, using  $\omega_m$ ,  $n_s$  and  $d_A(z_*)$  as determined by Planck together with the Riess et al. value of the Hubble parameter. This yields shifts in the background parameters by about  $1\sigma$ :

$$100 \Omega_K = +0.37 \pm 0.47, \quad (\text{for } w = -1) \quad (20)$$

$$w = -1.07 \pm 0.082, \quad (\text{for } \Omega_K = 0). \quad (21)$$

Hence taking this shift to the distance into account together with local observations, the CMB data remains consistent with a minimal flat  $\Lambda$ CDM model. This is in contrast to the standard calculation of distance where it is difficult to relieve the tension between local measurements of  $H_0$  [3–6] and the value from CMB observations.

Clearly, the analysis presented here is not definitive for several reasons. First, as mentioned above, the Planck measurements of cosmological parameters are not independent and especially the Planck value of  $\omega_m$  is not completely independent of the distance  $d_A$ . We have also assumed a simple linear transfer function [17]. A full likelihood analysis should be performed with  $\langle \Delta \rangle$  properly included.

Furthermore, aggregated lensing not only leads to a average shift in the distance to the CMB but  $\Delta$  is actually direction dependent [13]. We expect its fluctuations to be imprinted as additional fluctuations in the CMB. However, since the dominant contribution comes from very small scales, we expect them to show up mainly at high

$\ell > 2000$  and we believe that the effect on the mean distance discussed here is the dominant one in present CMB experiments. To do a fully consistent analysis, which combines second order fluctuations in the distance with temperature perturbations, a 3rd order Boltzmann solver would be needed. A interesting future project which is (far!) beyond the scope of the present paper.

We also note that higher-order contributions to  $\langle \Delta \rangle$  will be small, though ultimately necessary as observations improve. We anticipate they will be dominated by terms such as  $(\delta d_A)^4$ , which will lead to a percent level correction to our second-order correction. Thus, the main contribution to aggregated lensing is from (6).

#### IV. CONCLUSIONS

We have demonstrated an important overall shift in the distance redshift relation when the aggregate of all lensing events is considered, calculated by averaging over an ensemble of universes. This result is a consequence of flux conservation at second-order in perturbation theory. This is a purely relativistic effect with no Newtonian counterpart – and it is the first quantitative prediction for a significant change to the background cosmology when averaging over structure [21]. The extraordinary amplification of aggregated lensing comes mainly from the integrated lensing of structure on scales in the range  $1 - 100 \text{ Mpc}$ , although structure down to  $10 \text{ kpc}$  scales contributes significantly. We have estimated the size of the effect using a linear transfer function which slightly underestimates power on small scales at high redshift, so this provides a robust lower limit to the overall amplitude. Higher-order corrections from relativistic perturbation theory will enter  $\mathcal{O}((\delta d_A)^4)$ , making (6) the main contribution in general.

This isotropic shift is particularly important for high redshift, apparently giving a change to the distance to the CMB of one percent. What does this mean? We have argued that the shift can be interpreted as a change to the inferred background cosmology. Assuming that observations of the CMB really measures the area distance implies that fitting to the minimal  $\Lambda$ CDM model leads to an underestimation of the Hubble parameter by 5%. We have considered the consequences for analysing the CMB, and have argued that parameter estimation could be strongly affected – parameter constraints can be shifted by more than  $1\sigma$ . Because the shift  $\langle \Delta \rangle$  increases the distance relative to the background, the corrections to the background to compensate for this increase in distance are achieved by increasing  $h$ . In particular, we have shown that a higher  $h$  is naturally preferred over the low value found by Planck [3], in line with local data [4]. For current and future redshift surveys, the effective model yields sub-percent changes to the curvature, and dark energy parameters, which will eventually be important for precision cosmology of the coming decade.

Of course, our analysis is rather simplistic. We have

calculated the expected correction to the distance of an infinitesimal spot on the CMB sky and extrapolated up to sound-horizon scales. We have assumed that the CMB temperature map directly measures the distance to last scattering, when in fact it is inferred indirectly. The full correction to the background parameters corresponding to aggregated lensing – a change to the global distance from the sum of all lensing events – may well give quite different results. Nevertheless, as our simple consideration has led to 5% difference in the estimated Hubble parameter, and 10% in  $\Omega_m$ , it is clear that 2nd order aggregated lensing has to be included in a CMB calculation which aspires to a precision of 0.1%. Beyond linear perturbation theory, the correspondence between the expectation value of observables and the averaged background model becomes very subtle to define precisely. We have shown that it could lead to important corrections to the background when asking, What is the distance to

the CMB?

**Acknowledgements:** Special thanks to Camille Bonvin, Antony Lewis and Fabien Schmidt for extensive and very helpful discussions. We are also grateful for discussions with and comments from Phil Bull, George Ellis, Pedro Ferreira, Daniel Holz, Martin Kunz, Giovanni Marozzi, Uros Seljak, Bjoern Schaefer, Bob Wald and David Weinberg. CC and RM are supported by the South African National Research Foundation. OU and RM are supported by the South African Square Kilometre Array Project. RM acknowledges support from the UK Science & Technology Facilities Council (grant ST/K0090X/1). RD acknowledges support from the Swiss National Science Foundation.

## Appendix A: The derivation of (6) geodesic deviation

The dominant terms in the ensemble average come from those with the largest number of screen space derivatives in them. Since a pure divergence at second-order will give zero ensemble average (or monopole), we can present a simple derivation of the relevant term.

We consider a bundle of future-pointing light rays around a given ray with momentum  $k^a$ ,  $k^b \nabla_b k^a = 0$ ,  $k_a k^a = 0$ . A light ray arriving at the observer from a slightly different direction is given by its angular position at the observer,  $\alpha$ . The Jacobi map maps  $\alpha$ , to a screen space position  $x_\perp(\lambda)$ , where  $\lambda$  is the affine parameter along the geodesic  $k^a$ . As  $dx_\perp^a/d\lambda = x_\perp^b \nabla_b k^a$  (see e.g. [25], eq. (3.50)), the area on the screen space satisfies

$$\frac{dA}{d\lambda} = \nabla_a k^a A = \theta A. \quad (A1)$$

Where  $\theta = \nabla_a k^a$  is the rate of expansion of the null congruence. Defining the area distance

$$d_A \propto \sqrt{A}, \quad (A2)$$

we obtain the basic equation for the area distance, valid in any spacetime:

$$k^a \nabla_a d_A = \frac{1}{2} \theta d_A. \quad (A3)$$

This definition gives the proper area of a source,  $dA$ , in terms of the solid angle at the observer,  $d\Omega$ , as  $dA = d_A^2 d\Omega$ . We expand order-by-order on a perturbed Minkowski background where  $\bar{d}_A = \chi = \eta_0 - \eta$ ,  $\bar{\theta} = -2/\chi$ ,  $\bar{k}^a \nabla_a = -D/d\chi$ , as  $d_A = \bar{d}_A + \delta d_A + \frac{1}{2} \delta^2 d_A$ , similarly  $k^a + \delta k^a + \frac{1}{2} \delta^2 k^a$  and  $\theta = \bar{\theta} + \delta\theta + \frac{1}{2} \delta^2 \theta$ . Photon geodesics are conformally invariant, hence we can remove the irrelevant cosmic scale factor in this calculation. A background conformal transformation then gives the full area distance [14].

The first order perturbation of the area distance is well known, see, e.g. [7, 14]

$$\delta d_A(\chi) = \int_0^\chi d\tilde{\chi} (\tilde{\chi} - \chi) \tilde{\chi} \nabla_\perp^2 \Phi(\tilde{\chi}). \quad (A4)$$

Inserting this in (A3) and neglecting subdominant terms (at first-order the perturbation of  $\delta k^a$  contains at most one  $\nabla_\perp$ ), we obtain the perturbation of the null expansion

$$\delta\theta(\chi) = -2 \frac{d}{d\chi} \frac{\delta d_A}{\chi} = 2 \int_0^\chi d\tilde{\chi} \left( \frac{\tilde{\chi}}{\chi} \right)^2 \nabla_\perp^2 \Phi(\tilde{\chi}) \quad (A5)$$

At second-order,

$$-\frac{d}{d\chi} \delta^2 d_A + 2\delta k^a \nabla_a \delta d_A + \delta^2 k^a \nabla_a \chi = \frac{1}{2} \chi \delta^2 \theta + \delta d_A \delta\theta + \frac{1}{2} \bar{\theta} \delta^2 d_A \quad (A6)$$

The left hand side of this equation contains the relevant post-Born-approximation terms which reflect the fact that the geodesic deviates from a straight line. The right hand side is the impact of the change of the rate of expansion of the geodesic congruence on the distance. Now, we can neglect:  $\delta^2 k^a \nabla_a \chi = \delta^2 k_{\parallel} \nabla_{\parallel} \chi$  as it has only lower derivative terms. Then  $\delta^2 \theta \sim \nabla_{\perp i} \delta^2 k^i$  + lower derivative terms. Since the leading order is a 2-divergence it will not contribute to the mean.[28] This leaves

$$\frac{d\delta^2 d_A}{d\chi} - \frac{1}{\chi} \delta^2 d_A = -\delta d_A \delta \theta + 2\delta k^a \nabla_a \delta d_A \quad (\text{A7})$$

$$= -3\delta d_A \delta \theta = 6\delta d_A \frac{d}{d\chi} \frac{\delta d_A}{\chi} \quad (\text{A8})$$

where the second equality holds up to a pure divergence on writing  $\nabla_{\perp i}(\delta k^i \delta d_A) = \nabla_{\perp i}(\delta k^i) \delta d_A + \delta k^i \nabla_{\perp i}(\delta d_A)$  and we have used (A5) for the last equality. Noting that the right hand side of (A8) equals  $\chi \frac{d}{d\chi} (\frac{\delta^2 d_A}{\chi})$  we obtain

$$\frac{\delta^2 d_A(\chi_s)}{\chi_s} = 3 \left( \frac{\delta d_A(\chi_s)}{\chi_s} \right)^2 \quad (\text{A9})$$

$$\delta^2 d_A(\chi_s) = 3\chi_s \left[ \int_0^{\chi_s} d\chi \left( 1 - \frac{\chi}{\chi_s} \right) \chi \nabla_{\perp}^2 \Phi(\chi) \right]^2 \quad (\text{A10})$$

$$= 3\chi_s \left[ \int_0^{\chi_s} \frac{d\chi}{\chi} \left( 1 - \frac{\chi}{\chi_s} \right) \Delta_2 \Phi(\chi) \right]^2 \quad (\text{A11})$$

For the last equal sign we used  $\chi^2 \nabla_{\perp}^2 = \Delta_2$ .

This can be interpreted in a variety of different ways, on moving the angular derivatives outside the integrals, and expanding the pure divergence  $\nabla_{\perp i}[\nabla_{\perp j}(\nabla^{\perp i} X \nabla^{\perp j} Y)]$ . This can be used to re-write this in terms of the (linear null shear)<sup>2</sup> and the bending angle coupled to gradients in the linear convergence as presented in [13].

### Consistency with the Sachs equation

For completeness, we can show consistency with the Sachs equations, written in terms of the distance. We use the notation of Ref. [13].

$$\frac{d^2 d_A}{d\lambda^2} = -\frac{1}{2} [R_{ab} k^a k^b + \Sigma_{ab} \Sigma^{ab}] D_A \quad (\text{A12})$$

where the shear obeys

$$\frac{D\Sigma_{\langle ab \rangle}}{d\lambda} = -\theta \Sigma_{ab} + N_{\langle a}{}^e N_{b \rangle}{}^f R_{efcd} k^c k^d. \quad (\text{A13})$$

These are fully non-linear equations valid in any spacetime. The second-order part of (A12) is

$$\begin{aligned} \frac{d^2 \delta^2 d_A}{d\lambda^2} = & -2\delta k^a \delta k^b \nabla_a \nabla_b d_A - 2k^b \delta^2 k^a \nabla_a \nabla_b d_A - 4\delta k^a k^b \nabla_a \nabla_b \delta d_A - 2\delta k^a \nabla_a \delta k_b \nabla^b d_A \\ & - \frac{D\delta^2 k_a}{d\lambda} \nabla^a d_A - \delta^2 k^a \nabla_a k_b \nabla^b d_A - 2 \frac{D\delta k_a}{d\lambda} \nabla^a \delta d_A - 2\delta k^a \nabla_a k_b \nabla^b \delta d_A - \delta d_A \delta R_{ab} k^a k^b \\ & + d_A \left[ -\delta \Sigma_{ab} \delta \Sigma^{ab} - 2\delta k^a k^b \delta R_{ab} - \frac{1}{2} \delta^2 R_{ab} k^a k^b \right]. \end{aligned} \quad (\text{A14})$$

Neglecting contributions that do not have 4 derivatives this becomes

$$\frac{d^2 \delta^2 d_A}{d\chi^2} = -4\delta k^a k^b \nabla_a \nabla_b \delta d_A - 2k^b \nabla_b \delta k_a \nabla^a \delta d_A - 2\delta k^a \nabla_a k_b \nabla^b \delta d_A - \delta d_A \delta R_{ab} k^a k^b - \chi \delta \Sigma_{ab} \delta \Sigma^{ab}. \quad (\text{A15})$$

Next we can calculate, including only 4-derivative terms that do not form a divergence [ $' = d/d\chi$ ]:

$$\delta k^a k^b \nabla_a \nabla_b \delta d_A = \delta \theta \delta d_A' \quad (\text{A16})$$

$$k^b \nabla_b \delta k_a \nabla^a \delta d_A = \frac{1}{\chi} \delta \theta \delta d_A + \delta \theta' \delta d_A \quad (\text{A17})$$

$$\delta k^a \nabla_a k_b \nabla^b \delta d_A = \frac{1}{\chi} \delta \theta \delta d_A \quad (\text{A18})$$

$$\delta d_A \delta R_{ab} k^a k^b = -2\nabla_{\perp}^2 \Phi \delta d_A. \quad (\text{A19})$$



For the shear we solve (A13):

$$\delta\Sigma_{ij} = 2 \int_0^\chi d\tilde{\chi} \left( \frac{\tilde{\chi}}{\chi} \right)^2 \nabla_{\perp\langle i} \nabla_{\perp j \rangle} \Phi(\tilde{\chi}). \quad (\text{A20})$$

Now, noting that, up to a pure divergence,

$$\nabla_{\perp\langle i} \nabla_{\perp j \rangle} A \nabla^{\perp\langle i} \nabla^{\perp j \rangle} B = \frac{1}{2} \nabla_{\perp}^2 A \nabla_{\perp}^2 B, \quad (\text{A21})$$

we find

$$\delta\Sigma_{ij} \delta\Sigma^{ij} = \frac{1}{2} (\delta\theta^2) + \text{divergence}. \quad (\text{A22})$$

Inserting these results in (A15) we find

$$\frac{d^2 \delta^2 d_A}{d\chi^2} = -4 \left( \frac{d}{d\chi} + \frac{1}{\chi} \right) \delta\theta \delta d_A + 2 \nabla_{\perp}^2 \Phi \delta d_A - \frac{1}{2} \chi \delta\theta^2. \quad (\text{A23})$$

Next we use

$$\frac{d}{d\chi} \delta\theta = 2 \nabla_{\perp}^2 \Phi - \frac{2}{\chi} \delta\theta, \quad (\text{A24})$$

$$\frac{d}{d\chi} \delta d_A = \frac{1}{\chi} \delta d_A - \frac{1}{2} \chi \delta\theta; \quad (\text{A25})$$

these imply

$$2 \nabla_{\perp}^2 \Phi \delta d_A - \frac{1}{2} \chi \delta\theta^2 = \left( \frac{d}{d\chi} + \frac{1}{\chi} \right) \delta\theta \delta d_A \quad (\text{A26})$$

so that

$$\frac{d^2 \delta^2 d_A}{d\chi^2} = -3 \left( \frac{d}{d\chi} + \frac{1}{\chi} \right) \delta\theta \delta d_A, \quad (\text{A27})$$

which is the derivative of

$$\frac{d\delta^2 d_A}{d\chi} - \frac{1}{\chi} \delta^2 d_A = -3 \delta d_A \delta\theta, \quad (\text{A28})$$

in agreement with (A11).

## Appendix B: Integral formulation, and the meaning of linear terms at second-order

Here we present a more conventional integral formulation of the derivation of (6). We integrate (A3) directly along the perturbed geodesic:

$$2 \ln d_A(\chi_s, \mathbf{n}_{\text{obs}}) = \int_{\lambda_o}^{\lambda_s} d\lambda \theta \quad (\text{B1})$$

$$= - \int_{\lambda_o}^{\lambda_s} d\lambda \frac{2}{\chi} + \int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta + \int_{\lambda_o}^{\lambda_s} d\lambda \delta^2\theta \quad (\text{B2})$$

Note that the left hand side is a scalar evaluated at the *observed* source position  $S$ , as we integrate from observer to source. On the right hand side, we have expanded  $\theta$ , and have kept the integrals along the perturbed geodesic given by the curve  $x^a(\lambda) = \int_{\lambda_o}^{\lambda} d\tilde{\lambda} k^a(\tilde{\lambda})$ . The first term of (B2) is trivial to evaluate, and the last leads to a 2-divergence. The middle one, which is the integral of a first order quantity on a perturbed path which we need to second-order. Referring to Fig. 3, we can expand the integral about a background geodesic in different ways:

Let us consider expanding about geodesic 1.

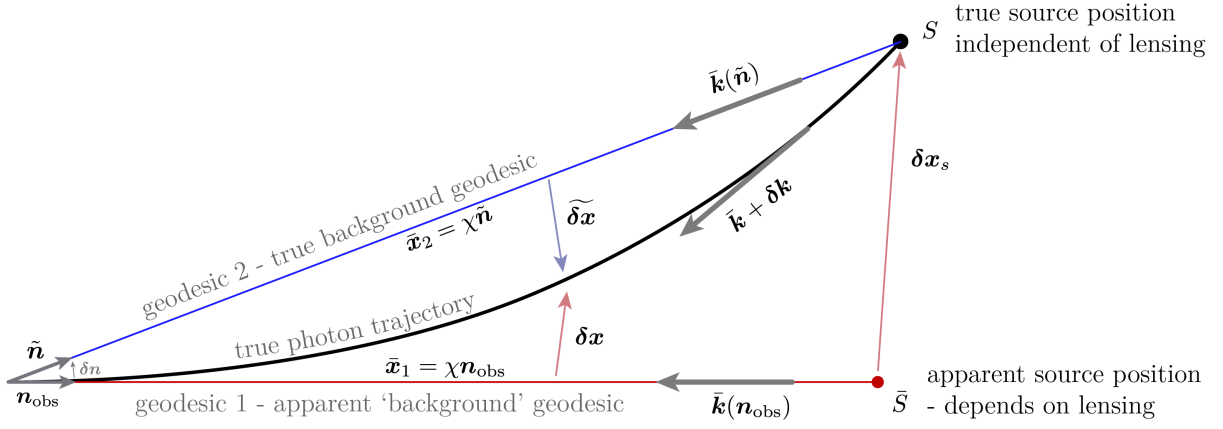


Figure 3: Real and apparent source positions relative to the observer give different ‘background’ geodesics:

**geodesic 1:** the geodesic from the observer in the same physical direction as the observed source,  $\mathbf{n}_{\text{obs}}$ . This straight line ends at a fictitious place for the source - the apparent position.

**geodesic 2:** the geodesic from the observer to the physical position to the observed source. This straight line starts at an unobservable place for the observer.

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = \int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}_1(\lambda)) + \int_{\lambda_o}^{\lambda_s} d\lambda \delta x^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B3})$$

We can write on any *background* geodesic

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}, \bar{x}(\lambda)) = - \int_0^{\chi_s} d\chi \delta\theta(\mathbf{n}, \bar{x}(\chi)) = 2 \frac{\delta d_A}{\chi}. \quad (\text{B4})$$

On the right hand side of this, we must specify the location the function is evaluated at, either in terms of a physical spacetime position, or a direction from the observer.

Then we can write (B3) as:

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = 2 \frac{\delta d_A}{\chi} \Big|_{\bar{x}_{\tilde{S}}} + \int_{\lambda_o}^{\lambda_s} d\lambda \delta x^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B5})$$

$$= 2 \frac{\delta d_A}{\chi} \Big|_{x_s} - 2 \delta x^i \nabla_i \frac{\delta d_A}{\chi} + \int_{\lambda_o}^{\lambda_s} d\lambda \delta x^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B6})$$

where the second line shifts the linear term evaluated at the image position  $\tilde{S}$  to the true source position  $S$ . In the second line we now have a result where both sides are evaluated at the same physical point, whereas the first line is an equation where the left hand side is evaluated at the true source position  $S$ , and the right hand side at the apparent position  $\tilde{S}$ . If we take an ensemble average at this point, the first term on the second line is linear, evaluated at the true source position, and can be ignored.

Up to a total divergence we then obtain

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = 2 \frac{\delta d_A}{\chi} \Big|_{x_s} - 2 \delta x^i \nabla_i \frac{\delta d_A}{\chi} + \int_{\lambda_o}^{\lambda_s} d\lambda \delta x^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B7})$$

$$= 2 \frac{\delta d_A}{\chi} \Big|_{x_s} + 2 (\nabla_i \delta x^i) \frac{\delta d_A}{\chi} + \int_0^{\chi_s} d\chi (\nabla_i \delta x^i) \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B8})$$

$$= 2 \frac{\delta d_A}{\chi} \Big|_{x_s} + 2 (\nabla_i \delta x^i) \frac{\delta d_A}{\chi} + (\nabla_i \delta x^i) \int_0^{\chi_s} d\chi \delta\theta - \int d\chi \left[ \int_0^{\chi} d\chi' \delta\theta(\chi') \right] \frac{d}{d\chi} \nabla_i \delta x^i \quad (\text{B9})$$

$$= 2 \frac{\delta d_A}{\chi} \Big|_{x_s} + 2 \int d\chi \frac{\delta d_A}{\chi} \frac{d}{d\chi} \nabla_i \delta x^i \quad (\text{B10})$$

Now, evaluating  $\nabla_i \delta x^i$ :

$$\frac{d}{d\chi} \nabla_i \delta x^i = \frac{d}{d\chi} (N^{ij} \nabla_i \delta x_j) = -\delta\theta + \text{subdominant terms} \quad (\text{B11})$$

which implies

$$\nabla_i \delta x^i = - \int_0^\chi d\chi \delta\theta = 2 \frac{\delta d_A}{\chi} + \text{subdominant terms.} \quad (\text{B12})$$

Ignoring subdominant terms we then have

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = 2 \frac{\delta d_A}{\chi} \Big|_{x_s} + 2 \left( \frac{\delta d_A}{\chi} \right)^2 \quad (\text{B13})$$

Note that if we had not employed the shift, the formula would read, from (B3),

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = 2 \frac{\delta d_A}{\chi} \Big|_{\bar{x}_s} - 2 \left( \frac{\delta d_A}{\chi} \right)^2 \quad (\text{B14})$$

Both formulas are correct, but they mean slightly different things.

To identify which we want to use in a given situation, let us expand the left hand side of (B2):

$$2 \ln d_A = 2 \ln \chi_s + 2 \frac{\delta d_A}{\chi_s} + \frac{\delta^2 d_A}{\chi_s} - \left( \frac{\delta d_A}{\chi_s} \right)^2 \quad (\text{B15})$$

The meaning of the second-order term changes depending on whether we choose to evaluate the linear term at the apparent position or the true position of the source. Now, if we choose to evaluate the linear term at the physical position of the source we have to use (B13):

$$+ 2 \frac{\delta d_A}{\chi_s} \Big|_{x_s} + \frac{\delta^2 d_A}{\chi_s} - \frac{\delta d_A^2}{\chi_s^2} = 2 \frac{\delta d_A}{\chi} \Big|_{x_s} + 2 \left( \frac{\delta d_A}{\chi_s} \right)^2 \quad (\text{B16})$$

the linear terms cancel, and so we derive:

$$\frac{\delta^2 d_A}{\chi_s} = 3 \left( \frac{\delta d_A}{\chi_s} \right)^2 \quad \text{if } \delta d_A \text{ is evaluated at the physical position } S, \quad (\text{B17})$$

in agreement with (A9)

On the other hand, if we choose to evaluate the terms in  $d_A$  at the background position of the source we would use (B14) to write:

$$+ 2 \frac{\delta d_A}{\chi_s} \Big|_{\bar{x}_s} + \frac{\delta^2 d_A}{\chi_s} - \frac{\delta d_A^2}{\chi_s^2} = 2 \frac{\delta d_A}{\chi} \Big|_{\bar{x}_s} - 2 \left( \frac{\delta d_A}{\chi_s} \right)^2 \quad (\text{B18})$$

from which we have:

$$\frac{\delta^2 d_A}{\chi_s} = - \left( \frac{\delta d_A}{\chi_s} \right)^2 \quad \text{if } \delta d_A \text{ is evaluated at the apparent position of the source} \quad (\text{B19})$$

However, the apparent position of the source is a perturbed quantity and a first order perturbation evaluated there contains a second order contribution such that its ensemble average is not guaranteed to vanish.

**Along geodesic 2:** We now repeat the calculation using geodesic 2 as our reference geodesic:

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = \int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\tilde{\mathbf{n}}, \bar{x}_2(\lambda)) + \int_{\lambda_o}^{\lambda_s} d\lambda \widetilde{\delta x}^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B20})$$

$$= 2 \frac{\delta d_A}{\chi} \Big|_{\tilde{\mathbf{n}}} + \int_{\lambda_o}^{\lambda_s} d\lambda \widetilde{\delta x}^i \nabla_i \delta\theta(\mathbf{n}_{\text{obs}}, \bar{x}(\lambda)) \quad (\text{B21})$$

In the first term we need the distinction between  $\tilde{\mathbf{n}}$  and  $\mathbf{n}_{\text{obs}}$ , in the second, which is already second order this is not necessary. The first term is a function evaluated in terms of an unobservable direction  $\tilde{\mathbf{n}}$  to the actual source location. To evaluate the second term we use the following relations which are follow from Fig 3

$$\bar{x}_1 + \delta x = \bar{x}_2 + \widetilde{\delta x} \quad (\text{B22})$$

$$\widetilde{\delta x} = \delta x - \chi \delta n = \delta x - \frac{\chi}{\chi_s} \delta x_s \quad (\text{B23})$$

where the last step follows by similar triangles. Substituting this into (B21) we find,

$$\int_{\lambda_o}^{\lambda_s} d\lambda \delta\theta(\mathbf{n}_{\text{obs}}, x(\lambda)) = 2 \frac{\delta d_A}{\chi} \bigg|_{\tilde{\mathbf{n}}} + 2 \left( \frac{\delta d_A}{\chi} \right)^2 \quad (\text{B24})$$

in agreement with (B13). Switching from  $\tilde{\mathbf{n}}$  to  $\mathbf{n}_{\text{obs}}$  reproduces (B14).

This derivation makes it clear that the formula in Appendix A applies *only when the linear term is evaluated at the physical position of the source*. When we evaluate ensemble averages this is the formula we need to use. The apparent position of the source moves in response to lensing events whereas the physical position does not. Thus, the term which ‘appears’ as linear when we use the observed apparent position of the source,  $\tilde{S}$ , in fact has second-order corrections which can move  $\tilde{S}$  around.

### Appendix C: Derivation of (6) from the lensing map

Let us start with the lens map [26],

$$\mathcal{D} = \lambda \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (\text{C1})$$

where  $\kappa$  denotes the convergence,  $\gamma = \gamma_1 + i\gamma_2$  is the complex shear and  $\lambda$  is the affine parameter of the photon geodesic. We perturb all quantities to 2nd order

$$\lambda = \chi + \lambda^{(1)} + \frac{1}{2} \lambda^{(2)} \quad (\text{C2})$$

$$\kappa(S) = \kappa^{(1)}(\bar{S}) + \frac{1}{2} \kappa^{(2)}(\bar{S}) \quad (\text{C3})$$

$$\gamma_{1,2}(S) = \gamma_{1,2}^{(1)}(\bar{S}) + \dots \quad (\text{C4})$$

Here we indicate how perturbed quantities are understood. Namely the true  $\kappa$  at the true source position  $S$  is approximated by a 1st and second order term both evaluated at the image position  $\bar{S}$ , see Fig. 3.

As we have argued before, perturbations of the affine parameter  $\lambda$  are subdominant. Up to second order we then have

$$d_A^2(S) = \det(\mathcal{D}) = \chi^2 \left[ 1 - 2\kappa^{(1)}(\bar{S}) - \kappa^{(2)}(\bar{S}) + (\kappa^{(1)})^2(\bar{S}) - (\gamma_1^{(1)})^2(\bar{S}) - (\gamma_2^{(1)})^2(\bar{S}) \right] = \chi^2 \left[ 1 - 2\kappa^{(1)}(\bar{S}) + t(\bar{S}) \right], \quad (\text{C5})$$

where  $t$  is a total divergence. This can be verified, e.g. using the second order lensing expressions which have been calculated in Ref. [26]. Let us take the square root:

$$d_A(S) = \chi \left[ 1 - \kappa^{(1)}(\bar{S}) - \frac{1}{2} \left( \kappa^{(1)}(\bar{S}) \right)^2 + \frac{1}{2} t(\bar{S}) \right]. \quad (\text{C6})$$

We want to determine the 2nd order contribution to this quantity. For this we subtract the 0th and 1st order. But to subtract the 1st order we have to evaluate both sides at the same position. To 2nd order this does make a difference.

$$d_A(S) = \chi \left[ 1 + \frac{\delta d_A(S)}{\chi} + \frac{1}{2} \frac{\delta^2 d_A(S)}{\chi} \right] \quad (\text{C7})$$

Expanding also

$$\kappa^{(1)}(\bar{S}) = \kappa^{(1)}(S) - \delta x^a \nabla_a \kappa^{(1)}(\bar{S}) = \kappa^{(1)}(S) + (\nabla_a \delta x^a) \kappa^{(1)}(\bar{S}) = \kappa^{(1)}(S) - 2[\kappa^{(1)}(\bar{S})]^2 + \text{total divergence} \quad (\text{C8})$$

Here we have used  $(\nabla_a \delta x^a) = -2\kappa^{(1)}$  to first order.

Now we evaluate both sides at the same position and can safely subtract the 0th and 1st order terms. With the first order identity

$$-\kappa^{(1)}(S) = \frac{\delta d_A(S)}{\chi}$$

we finally obtain

$$\frac{\delta^2 d_A(S)}{\chi} = 3(\kappa^{(1)})^2(\bar{S}) + \text{total divergence} = 3 \frac{(\delta d_A)^2}{\chi^2} + \text{total divergence}. \quad (\text{C9})$$

### Appendix D: Proof that the area of the CMB is larger with structure present.

The area of a surface of constant redshift is the integral of  $d_A^2$  over the observers sky. Let us consider  $d_A^2$ . In general this obeys:

$$k^a \nabla_a d_A^2 = \theta d_A^2. \quad (D1)$$

Written relative to the distance in the background we can write this as (to all orders in perturbations)

$$k^a \nabla_a \left( \frac{d_A}{\chi} \right)^2 = \left( \theta - \frac{2}{\chi} k^a \nabla_a \chi \right) \left( \frac{d_A}{\chi} \right)^2. \quad (D2)$$

Now let us expand the derivative into a background and total perturbation part:

$$k^a \nabla_a = -\frac{d}{d\chi} + b \Delta k^a \nabla_a. \quad (D3)$$

We have made no approximation here, we just write  $k^a = \bar{k}^a + \Delta k^a$ . The  $b$  is just a constant to keep track of the post-Born terms,  $b = 1$  includes them,  $b = 0$  just integrates on the background geodesic. Setting

$$\left( \frac{d_A}{\chi} \right)^2 = 1 + \frac{\Delta d_A^2}{\chi^2} \quad \text{and} \quad \Delta \theta = \theta + \frac{2}{\chi}. \quad (D4)$$

we can write

$$\frac{d}{d\chi} \frac{\Delta d_A^2}{\chi^2} = -\Delta \theta \left( 1 + \frac{\Delta d_A^2}{\chi^2} \right) + b \Delta k^a \nabla_a \frac{\Delta d_A^2}{\chi^2}. \quad (D5)$$

So far, these are just definitions, not yet implying approximations (though all  $\Delta$  quantities vanish on the background). Now we can start applying approximations used above. Up to subdominant terms  $\nabla_a \Delta k^a = \Delta \theta$ . Therefore up to a total divergence and subdominant terms,

$$\Delta k^a \nabla_a \frac{\Delta d_A^2}{\chi^2} = -\Delta \theta \frac{\Delta d_A^2}{\chi^2} \quad (D6)$$

neglecting lower derivative terms from the perturbed Christoffel symbols. With this

$$\frac{d}{d\chi} \frac{\Delta d_A^2}{\chi^2} + (1+b) \Delta \theta \frac{\Delta d_A^2}{\chi^2} = -\Delta \theta. \quad (D7)$$

Integrating,

$$\frac{\Delta d_A^2}{\chi^2} = -e^{-(1+b) \int^\chi d\chi' \Delta \theta} \int d\chi' \Delta \theta e^{(1+b) \int^{\chi'} d\chi'' \Delta \theta}. \quad (D8)$$

Integrating this result by parts, we find

$$\frac{\Delta d_A^2}{\chi^2} = -\int_0^\chi d\chi' \Delta \theta + (1+b) e^{-(1+b) \int^\chi d\chi' \Delta \theta} \int d\chi' \left( \int^{\chi'} d\chi'' \Delta \theta \right) \Delta \theta e^{(1+b) \int^{\chi'} d\chi'' \Delta \theta}. \quad (D9)$$

The exponential terms only affect the solution at 3rd order, so we can ignore them. Using  $\int d\chi' \left( \int^{\chi'} d\chi'' \Delta \theta \right) \Delta \theta = \frac{1}{2} \left( \int_0^\chi d\chi' \Delta \theta \right)^2$ , we arrive at

$$\frac{\Delta d_A^2}{\chi^2} = -\int_0^\chi d\chi' \Delta \theta + \frac{1+b}{2} \left( \int_0^\chi d\chi' \Delta \theta \right)^2. \quad (D10)$$

The first term always forms a total divergence order-by-order, while the second term *always gives a positive contribution to the area*.

### Appendix E: Derivation of (10)

We start with (6):

$$\Delta_{\text{int}}(\chi_s, \mathbf{n}) = \frac{3}{2} \int_0^{\chi_s} d\chi_1 \frac{(\chi_s - \chi_1)}{\chi_s \chi_1} \int_0^{\chi_s} d\chi_2 \frac{(\chi_s - \chi_2)}{\chi_s \chi_2} \Delta_2 \Phi(\chi_1) \Delta_2 \Phi(\chi_2) \quad (\text{E1})$$

Expanding  $\Phi$  in spherical harmonics with (7), taking an ensemble average and performing one  $k$ -integral gives

$$\begin{aligned} \langle \Delta_{\text{int}} \rangle &= 6\pi \int_0^{\chi_s} d\chi_1 \frac{(\chi_s - \chi_1)}{\chi_s \chi_1} \int_0^{\chi_s} d\chi_2 \frac{(\chi_s - \chi_2)}{\chi_s \chi_2} g(\chi_1) g(\chi_2) \sum_{\ell_1 m_1} \sum_{\ell_2 m_2} i^{\ell_1 - \ell_2} \int \frac{dk}{k} \mathcal{P}_0 T(k)^2 j_{\ell_1}(k\chi_1) j_{\ell_2}(k\chi_2) \\ &\quad \times \int d\Omega_K Y_{\ell_1 m_1}(\hat{\mathbf{k}}) Y_{\ell_2 m_2}^*(\hat{\mathbf{k}}) \ell_1(\ell_1 + 1) \ell_2(\ell_2 + 1) Y_{\ell_1 m_1}(\mathbf{n}) Y_{\ell_2 m_2}^*(\mathbf{n}). \end{aligned} \quad (\text{E2})$$

Now we can perform the angular integral in  $k$ -space and use  $\int d\Omega_k Y_{\ell_1 m_1}(\hat{\mathbf{k}}) Y_{\ell_2 m_2}^*(\hat{\mathbf{k}}) = \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}$ , and from  $\sum_{m=-\ell}^{m=\ell} Y_{\ell m}(\mathbf{n}) Y_{\ell m}^*(\mathbf{n}) = (2\ell + 1)/4\pi$  we derive

$$\langle \Delta_{\text{int}} \rangle = \frac{3}{2} \int_0^{\chi_s} d\chi_1 \frac{(\chi_s - \chi_1)}{\chi_s \chi_1} \int_0^{\chi_s} d\chi_2 \frac{(\chi_s - \chi_2)}{\chi_s \chi_2} g(\chi_1) g(\chi_2) \sum_{\ell=0}^{\infty} (2\ell + 1) [\ell(\ell + 1)]^2 \int \frac{dk}{k} \mathcal{P}_0 T(k)^2 j_{\ell}(k\chi_1) j_{\ell}(k\chi_2).$$

This can be accurately simplified using the Limber approximation:

$$\int \frac{dk}{k} \mathcal{P}_0 T(k)^2 j_{\ell}(k\chi_1) j_{\ell}(k\chi_2) \simeq \frac{4\pi\chi_1}{(2\ell + 1)^3} \mathcal{P}_0 T^2 \left( k = \frac{\ell + 1/2}{\chi_1} \right) \delta(\chi_1 - \chi_2), \quad (\text{E3})$$

so

$$\langle \Delta_{\text{int}} \rangle = 6\pi \sum_{\ell=0}^{\infty} \left[ \frac{\ell(\ell + 1)}{2\ell + 1} \right]^2 \int_0^{\chi_s} d\chi \frac{(\chi_s - \chi)^2}{\chi \chi_s^2} g(\chi)^2 \mathcal{P}_0 T^2 \left( k = \frac{\ell + 1/2}{\chi} \right). \quad (\text{E4})$$

This is (10).

### Appendix F: Local terms at low redshift

For our purposes the most important local terms are the ones with the highest even numbers of spatial derivatives (odd numbers of spatial derivatives do not contribute to an ensemble average):

$$\begin{aligned} \Delta_{\text{loc}} &= \left( \frac{\mathcal{H}'}{2\mathcal{H}^2} - \frac{1}{\chi_s \mathcal{H}_s} - \frac{1}{2} \right) \nabla_{\parallel} v_s \nabla_{\parallel} v_s + \chi_s \nabla_{\parallel} \Phi_s \nabla_{\parallel} v_s \\ &\quad + \left( 1 - \frac{1}{\chi_s \mathcal{H}_s} \right) [\chi_s \nabla_{\parallel} v_s \nabla_{\parallel} v_s' + \nabla_{\perp i} v_s \nabla_{\perp}^i v_s + \chi_s \Phi_s \nabla_{\parallel}^2 v_s]. \end{aligned} \quad (\text{F1})$$

Here  $v_s$  is the velocity potential at the source, and  $\nabla_{\parallel}, \nabla_{\perp i}$  are the radial and transverse (screen-space) parts of the spatial derivative. This contains a mixture of radial and transverse velocity terms, as well as the redshift-space distortion effect.

For  $\langle \Delta_{\text{loc}} \rangle$ , all terms in (F1) are proportional to  $(\nabla_{\parallel} \Phi_0)^2$ , so  $\langle \Delta_{\text{loc}} \rangle \propto \langle (\nabla_{\parallel} \Phi_0)^2 \rangle$ , where

$$\langle (\nabla_{\parallel} \Phi_0)^2 \rangle = \frac{1}{2} \langle \nabla_{\perp i} \Phi_0 \nabla_{\perp}^i \Phi_0 \rangle = \frac{1}{3} \int dk k \mathcal{P}_0(k) T^2(k). \quad (\text{F2})$$

Summing up the terms in (F1) results in a small negative contribution, and is the dominant contribution at low- $z$  ( $z \lesssim 0.5$ ), as shown in Fig. 1.

- 034 [arXiv:1104.2933 [astro-ph.CO]].
- [3] P. A. R. Ade *et al.* [Planck Collaboration], arXiv:1303.5076.
  - [4] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha and W. Li *et al.*, *Astrophys. J.* **730**, 119 (2011) [Erratum-ibid. **732**, 129 (2011)] [arXiv:1103.2976 [astro-ph.CO]].
  - [5] V. Marra, L. Amendola, I. Sawicki and W. Valkenburg, *Phys. Rev. Lett.* **110**, no. 24, 241305 (2013) [arXiv:1303.3121 [astro-ph.CO]].
  - [6] V. C. Busti, C. Clarkson and M. Seikel, arXiv:1402.5429 [astro-ph.CO].
  - [7] C. Bonvin, R. Durrer and M. A. Gasparini, *Phys. Rev. D* **73**, 023523 (2006) [Erratum-ibid. **D 85**, 029901 (2012)] [astro-ph/0511183].
  - [8] I. Ben-Dayan, R. Durrer, G. Marozzi and D. J. Schwarz, arXiv:1401.7973.
  - [9] E. Barausse, S. Matarrese and A. Riotto, *Phys. Rev. D* **71**, 063537 (2005) [astro-ph/0501152].
  - [10] C. Clarkson and O. Umeh, *Class. Quant. Grav.* **28**, 164010 (2011) [arXiv:1105.1886].
  - [11] I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier and G. Veneziano, *Phys. Rev. Lett.* **110**, 021301 (2013) [arXiv:1207.1286].
  - [12] I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier and G. Veneziano, *JCAP* **1306**, 002 (2013) [arXiv:1302.0740].
  - [13] O. Umeh, C. Clarkson and R. Maartens, arXiv:1207.2109v2.
  - [14] O. Umeh, C. Clarkson and R. Maartens, arXiv:1402.1933.
  - [15] G. Fanizza, M. Gasperini, G. Marozzi and G. Veneziano, *JCAP* **1311** (2013) 019 [arXiv:1308.4935 [astro-ph.CO]].
  - [16] G. Marozzi, arXiv:1406.1135 [astro-ph.CO].
  - [17] D. J. Eisenstein and W. Hu, *Astrophys. J.* **496**, 605 (1998) [astro-ph/9709112].
  - [18] C. Clarkson, M. Cortes and B. A. Bassett, *JCAP* **0708**, 011 (2007) [astro-ph/0702670].
  - [19] R. E. Smith *et al.* [Virgo Consortium Collaboration], *Mon. Not. Roy. Astron. Soc.* **341**, 1311 (2003) [astro-ph/0207664].
  - [20] C. Clarkson, G. F. R. Ellis, A. Faltenbacher, R. Maartens, O. Umeh and J. -P. Uzan, *Mon. Not. Roy. Astron. Soc.* **426**, 1121 (2012) [arXiv:1109.2484].
  - [21] C. Clarkson, G. Ellis, J. Larena and O. Umeh, *Rept. Prog. Phys.* **74**, 112901 (2011) [arXiv:1109.2314].
  - [22] M. Vonlanthen, S. Rasanen and R. Durrer, *JCAP* **1008** (2010) 023 [arXiv:1003.0810].
  - [23] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha and W. Li *et al.*, *Astrophys. J.* **730**, 119 (2011) [Erratum-ibid. **732**, 129 (2011)]. [arXiv:1103.2976].
  - [24] D. N. Spergel *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **170**, 377 (2007) [astro-ph/0603449].
  - [25] P. Schneider, J. Ehlers and E.E. Falco, *Gravitational Lenses*, Springer Verlag (Berlin, 1992).
  - [26] F. Bernardeau, C. Bonvin and F. Vernizzi, *Phys. Rev. D* **81**, 083002 (2010). [arXiv:0911.2244]
  - [27] While it is strictly inaccurate to modify the background with curvature given our calculation of  $\langle \Delta \rangle$ , the errors will be  $\mathcal{O}(\Omega_K \Delta)$  which is small.
  - [28] We can see this once we have expanded in spherical harmonics we find  $\sum_{m=-\ell}^{m=\ell} Y_{\ell m}(\mathbf{n}) Y_{\ell m}^*(\mathbf{n}) = (2\ell + 1)/4\pi$  implies  $\sum_{m=-\ell}^{m=\ell} \nabla_{\Omega}^i Y_{\ell m}(\mathbf{n}) Y_{\ell m}^*(\mathbf{n}) = 0$ , and the divergence of this is zero too.